Proceedings 28th IEEE International Symposium on Multiple Valued Logic, Japan, 1998, pp.257-262

Functional Entropy and Decision Trees

Vasily Cheushev
Dept. of Programming Technology,
State University, 4 Skorina Av.,
200050, Minsk, Belarus,
e-mail:sasha@cacedu.unibel.by

Vlad Shmerko
Inst. of Computer Science,
Technical Univ., 49 Zolnierska St.,
71210 Szczecin, Poland

Dan A. Simovici
Dept. of Mathematics and Comp.Sci.
Univ. of Massachusetts at Boston
Boston, MA 02125, USA

Svetlana Yanushkevich
Inst. of Computer Science,
Technical Univ., 49 Zolnierska St.,
71210 Szczecin, Poland

Abstract

We introduce a technique to compute several information estimations for Boolean and multivalued functions. Special features of these estimations for completely and incompletely specified logic functions, including symmetric logic functions are investigated. Finally, we give an algorithm for determining various information measures for logical functions based on decision trees.

Keywords: information theoretic approach, information estimations, entropy, logic design, Boolean function, multivalued logic (MVL) function.

1 Introduction

Let $E_m = \{0, \dots, m-1\}$ be the set of logical values. Unless otherwise specified we assume that all logarithms are in base m. If M is a set, we denote by $\mathbf{Seq}(M)$ the set of sequences of elements of M. The length of a sequence $q \in \mathbf{Seq}(M)$ will be denoted by $|q|$ and the set of sequences of length k in $\mathbf{Seq}(M)$ will be denoted by $\mathbf{Seq}_k(M)$.

Let $D_0, \dots, D_{n-1}$ be n sets and let $u = (u_0, \dots, u_{n-1}) \in\in D_0 \times \cdots \times D_{n-1}$ be an n-tuple. Its restriction $(u_{i_0}, \cdots, u_{i_{k-1}})$ will be denoted by $u[i_0 \cdots i_{k-1}]$.

In this note the term *function* refers to partial functions, unless specified otherwise. The domain and range of a function f will be denoted by Dom (f) and Ran (f), respectively.

An m-ary U*-ordered tree* is a function $\mathtt{T} : \mathbf{Seq}(E_m) \longrightarrow U$ whose domain is a prefix-closed, finite subset of $\mathbf{Seq}(E_m)$. We refer to the elements of Dom (T) as the *nodes* of T. A leaf of T is a node q of T such q is a not a proper prefix of any other node of T. The nodes of an m-ary U-ordered tree are partitioned into *levels*. The j-th level of T is $\mathtt{T}_{[j]} = \{q \in \text{Dom (T)} \mid |q| = j\}$.

An information theoretic approach to solve logic design problems was tried for more than two decades. Typical issues that received attention include estimations for complexity of logic functions and characteristics of the interaction between values of functions and their arguments. In other cases, the information theoretic approach provided an alternative to traditional methods and approaches, in particular, when solving logic tasks through decision trees. Some of the earlier work [3, 5] deals with the conversion of decision tables to decision trees. Goodman and Smyth [4] proposed a general top-down mutual information algorithm to design decision trees from deterministic decision tables. A natural step is to use this result for minimization of logic functions. Because the truth table of logic function is a special case of decision table, the approach presented in [5] can be applied to the logic function minimization problem. It was shown in [6] that a minimal sum-of-products formula for a Boolean function can be obtained by entropy evaluation. In [8, 9] the results achieved in [6] are generalized to MVL functions. A notably successful application of the information-theoretical approach are the results on the relationship between entropy and the average power consumption of the circuits generated from Boolean functions (see [2]). The entropy measure was shown to provide an effective power estimate for single-output, multiple-output Boolean functions. A method to determine the entropy of large logic circuits was proposed in [7]. In [14] an information theoretic approach for the generation of near-optimum sequential fault location experiments is proposed. In [12, 1] it is shown that, unlike the notation of entropy of a probability distribution, the entropy of function can be regarded as an algebraic rather than a probabilistic concept. It means that a function (including a logic one) have so called functional entropy as a numerical characteristic, satisfying the group of axioms.

References [13] and [10] provide further interesting applications of information-theoretical approach.

The techniques proposed in this paper for computing information estimations for logic functions include rules to compute entropy and information quantity for completely and incompletely specified logic function, and for symmetric functions. Most examples are formulated, for simplicity, using binary functions; however, the algorithms presented have universal applicability to Boolean and multi-valued functions.

2 Information Measures of Functions

We assume that all combinations of values of variables occur with equal probability. A value of a function that occurs with the probability p carries a quantity of information equal to $-\log p$, where p is the probability of that value will occur.

The dependence of a function on an arbitrary variable can be expressed by the notion of *mutual information*.

The information carried by the value a of x_i is $I(x_i = a) = -\log p$, where p is the quotient between the number of tuples whose i-th components equal a and the total number of tuples. Similarly, the information carried by a value b of f is $I(f = b) = -\log q$, where q is the quotient between the number of tuples in the domain of f and the number of tuples for which f takes the value b.

Example 2.1 *If $f : E_m^n \longrightarrow E_m$ is the function defined by*

⋮					
x_i	0	1	0	1	0
⋮					
f	0	0	0	1	0

then the information carried by the values of x_i is $I(x_i = 0) = -\log_2 3/5 = 0.737$ bit, $I(x_i = 1) = -\log_2 2/5 = 1.322$ bit; the information carried $I(f = 0) = -\log_2 4/5 = 0.322$ bit and $I(f = 1) = -\log_2 1/5 = 2.322$ bit.

Recall that the conditional probability of a value b of a function f knowing the input value a of x_i is

$$p(f = b|x_i = a) = \frac{p((f = b) \land (x_i = a))}{p(x_i = a)}.$$

Similarly, the conditional probability of a value a of x_i given the value b of the function is

$$p(x_i = a|f = b) = \frac{p((f = b) \land (x_i = a))}{p(f = b)}.$$

Example 2.2 *For the partial Boolean function considered in Example 2.1 we have*

$$p(f = 0 \land x_i = 0) = 3/5, \quad p(f = 0 \land x_i = 1) = 1/5$$
$$p(f = 1 \land x_i = 0) = 0, \quad p(f = 1 \land x_i = 1) = 1/5.$$

Then,

$$\begin{aligned} p(f = 0|x_i = 0) &= (3/5)/(3/5) = 1, \\ p(f = 0|x_i = 1) &= (1/5)/(2/5) = 1/2 \\ p(f = 1|x_i = 0) &= 0, \\ p(f = 1|x_i = 1) &= (1/5)/(2/5) = 1/2. \end{aligned}$$

Definition 2.3 *The* relative information of the value b of the function given the value a_i of the input variable x_i *is* $I(f = b|x_i = a) = -\log_m p(f = b|x_i = a)$.

The relative information of the value a_i of the input variable x_i given the value b of the function *is* $I(x_i = a|f = b) = -\log_m p(x_i = a|f = b)$.

Example 2.4 *For the function considered in Example 2.1 we have*

$$\begin{aligned} I(f = 0|x_i = 0) &= -\log_2 1 = 0 \\ I(f = 0|x_i = 1) &= -\log_2 1/2 = 1 \\ I(f = 1|x_i = 0) &= 0 \\ I(f = 1|x_i = 1) &= -\log_2 1/2 = 1. \end{aligned}$$

Note that when the probability equals 0 *we suppose that the relative information is equal to* 0.

Definition 2.5 *The* mutual information between the value b of the function and the value a of the input variable x_i *is*

$$\begin{aligned} I(f = b; x_i = a) &= I(f = b) - I(f = b|x_i = a) \\ &= \log \frac{p((f = b) \land (x_i = a))}{p(x_i = a)p(f = b)}. \end{aligned}$$

The mutual information between the input variable x_i and the function f is

$$\begin{aligned} I(f; x_i) &= \sum_k \sum_\ell p((f = b_k) \land (x_i = a_\ell)) I(f = b_k; x_i = a_\ell) \\ &= \sum_k \sum_\ell p((f = b_k) \land (x_i = a_\ell)) \\ &\quad \cdot \log \frac{p((f = b_k) \land (x_i = a_\ell))}{p(x_i = a_\ell)p(f = b_k)}. \end{aligned}$$

Example 2.6 *For the function considered above we obtain* $I(f, x_i) = 0.6 \cdot 0.322 - 0.2 \cdot 0.678 + 0 + 0.2 \cdot 1.322 = 0.322$ *bit.*

The mutual information between an input variable x_i and a function f is used as to measure the dependence of the function f on the values of the variable x_i and vice-versa. In the above example, it means that x_i carries 0.322 bits of information about the function f.

Shannon's entropy of the variable x_i is $H(x_i) = \sum_{\ell=0}^{m-1} p(x_i = a_\ell) \log p(x_i = a_\ell)$, where m is the number of distinct values assumed by x_i; similarly, Shannon's entropy of the function f is $H(f) = \sum_{k=0}^{n-1} p(f = b_k) \log p(f = b_k)$, where n is the number of distinct values assumed by f.

Example 2.7 *For the function specified in Example 2.1 we have*

$$H(x_i) = -0.6 \cdot \log_2 0.6 - 0.4 \cdot \log_2 0.4 = 0.971 \textit{ bit}.$$

The entropy of the function is

$$H(f) = -0.8 \cdot \log_2 0.8 - 0.2 \cdot \log_2 0.2 = 0.722 \textit{ bit}.$$

Proposition 2.8 *The following statements hold:*

1. *For any variable x_i we have $0 \leq H(x_i) \leq 1$; similarly, for any function f, we have $0 \leq H(f) \leq 1$.*
2. *The entropy of any variable in a total function is 1.*
3. *The entropy of a constant function is 0.*

Definition 2.9 *The* conditional entropy of a function *f given a value a_i of the variable x_i is*

$$\begin{aligned} H(f|a_i) &= -\sum_{i=0}^{m-1} p((f = b_i) \wedge (x = a_i)) I(f = b_i | x_i = a_i) \\ &= -\sum_{i=0}^{m-1} p((f = b_i) \wedge (x = a_i)) \\ &\quad \cdot \log_m \frac{p((f = b) \wedge (x_i = a))}{p(x_i = a)}. \end{aligned}$$

The conditional entropy of a function f given a variable *x_i is*

$$\begin{aligned} H(f|x_i) &= -\sum_{j=0}^{m-1} \sum_{k=0}^{m-1} p((f = b_j) \wedge (x_i = a_k)) \\ &\quad \cdot \log_m p((f = b_j)|(x_i = a_k)). \end{aligned}$$

The joint entropy *of the function f and the variable x_i is*

$$\begin{aligned} H(f x_i) &= -\sum_{j=0}^{m-1} \sum_{k=0}^{m-1} p((f = b_j) \wedge (x_i = a_k)) \\ &\quad \cdot \log_m p((f = b_j) \wedge (x_i = a_k)). \end{aligned}$$

An elementary computation yields $H(f x_i) = H(f|x_i) + H(x_i) = H(x_i|f) + H(f)$.

Example 2.10 *Estimations for entropies and mutual information for binary functions are given below:*

	Entropy			Mutual Info.	
	$H(f)$	$H(f\|x_0)$	$H(f\|x_1)$	$I(f;x_0)$	$I(f;x_1)$
const0	0	0	0	0	0
$x_0 x_1$	0.81	0.5	0.5	0.31	0.31
$x_0 \bar{x}_1$	0.81	0.5	0.5	0.31	0.31
x_0	1	0	1	1	0
$\bar{x}_0 x_1$	0.81	0.5	0.5	0.31	0.31
x_1	1	0.5	0	0	1
$x_0 \ominus x_2$	1	1	1	0	0
$x_0 + x_1$	0.81	0.5	0.5	0.31	0.31
$x_0 \uparrow x_1$	0.81	0.5	0.5	0.31	0.31
$x_0 \sim x_1$	1	1	1	0	0
$\bar{x}_1$	1	1	0	0	1
$x_0 \rightarrow x_1$	0.81	0.5	0.5	0.31	0.31
$\bar{x}_0$	1	0	1	1	0
$x_1 \rightarrow x_0$	0.81	0.5	0.5	0.31	0.31
$x_0 \mid x_1$	0.81	0.5	0.5	0.31	0.31
const1	0	0	0	0	0

Note that dual Boolean functions have similar entropic characteristics.

Similar estimations are given below for multivalued functions:

f	$H(f)$	$H(fx)$	$H(f\|x)$	$H(x\|f)$	$I(f;x)$
$(x+y)_{mod\ 3}$	1	2	1	1	0
$(x+y)_{mod\ 4}$	1	2	1	1	0
$(xy)_{mod\ 3}$	0.906	1.665	0.665	0.759	0.241
$(xy)_{mod\ 4}$	0.876	1.626	0.626	0.750	0.250
$(xy)_{mod\ 5}$	0.957	1.800	0.800	0.843	0.157

Note that uncertainty of variables is maximal for total functions; in this case, the mutual information of variables is 0. Also, $H(x_i|x_j) = 1$ and $I(x_i|x_j) = 0$, for any such variables.

Example 2.11 *Consider the binary, n-argument symmetric functions f^n_{and} and f^n_{or} that give the conjunction and disjunctions of their arguments, respectively. For f^n_{and}, the probability associated with 0 is $p(f^n_{and} = 0) = (2^n - 1)/2^n$, while $p(f^n_{and} = 1) = 1/2^n$. Similarly, for f^n_{or} we have $p(f^n_{or} = 0) = 1/2^n$ and $p(f^n_{or} = 1) = (2^n - 1)/2^n$. Thus,*

$$H(f^n_{and}) = \frac{n}{2^n} - (1 - \frac{1}{2^n}) \log(1 - \frac{1}{2^n}),$$

and an equal value is obtained for $H(f^n_{or})$.

The value of the joint entropy of a symmetric function and a variable does not depend on any particular variable. Note that

$$\begin{aligned} p(f^n_{and} = 0 \wedge x = 0) &= \frac{1}{2} \\ p(f^n_{and} = 0 \wedge x = 1) &= \frac{1}{2} - \frac{1}{2^n} \\ p(f^n_{and} = 1 \wedge x = 0) &= 0 \\ p(f^n_{and} = 1 \wedge x = 1) &= \frac{1}{2^n}, \end{aligned}$$

which means that

$$H(f^n_{and}x) = \frac{1}{2} + \frac{n}{2^n} - (\frac{1}{2} - \frac{1}{2^n})\log(\frac{1}{2} - \frac{1}{2^n}).$$

Thus, we can compute the conditional entropy $H(f^n_{and}|x)$ and the mutual information $I(f^n_{and}; x)$ as

$$\begin{aligned} H(f^n_{and}|x) &= H(f^n_{and}x) - H(x) \\ I(f^n_{and}; x) &= H(f^n_{and}) - H(f^n_{and}|x). \end{aligned}$$

3 Mutual Information and Entropy for Functions and Sequences of Arguments

Let $f : E^n_m \longrightarrow E_m$ be an n-ary function and let $(x_{i_1}, x_{i_2}, \ldots, x_{i_q})$ be a sequence of arguments of this function. To simplify our notations we will denote the event $(f = b)$ simply by b; similarly, $(x_{i_k} = a_{i_k})$ will be denoted by a_{i_k}. The conjunction of two events $E \wedge F$ will be denoted by EF.

Theorem 3.1 *The joint entropy of the function f and the sequence of distinct variables $(x_{i_0}, \ldots, x_{i_{q-1}})$, $H(fx_{i_0} \cdots x_{i_{q-1}})$ is given by*

$$\begin{aligned} &H(fx_{i_0} \cdots x_{i_{q-1}}) \\ &= H(f) + H(x_{i_0}|f) + \cdots + H(x_{i_{q-1}}|fx_{i_0} \cdots x_{i_{q-2}}), \\ &= H(x_{i_{q-1}}) + H(x_{i_{q-2}}|x_{i_{q-1}}) + \\ &\quad \cdots + H(x_{i_0}|x_{i_1} \cdots x_{i_{q-1}}) + H(f|x_{i_{q-1}} \cdots x_{i_0}). \end{aligned}$$

The conditional entropy $H(f|x_{i_0} \cdots x_{i_{q-1}})$ is given by

$$H(f|x_{i_0} \cdots x_{i_{q-1}}) = H(fx_{i_0} \cdots x_{i_{q-1}}) - H(x_{i_0} \cdots x_{i_{q-1}}).$$

The mutual information between a function and a variable x_i can be expressed in entropy terms by $I(f; x_i) = H(f) - H(f|x_i)$. Similarly, we have:

$$I(f; x_i) = H(x_i) - H(x_i|f). \quad (1)$$

Next, this result is extended to a sequences of variables.

Theorem 3.2 *The mutual information between a sequence of distinct variables $(x_{i_0}, \ldots, x_{i_{q-1}})$ and the function f is given by $I(f; x_{i_0} \cdots x_{i_{q-1}}) = H(f) + H(x_{i_0} \cdots x_{i_{q-1}}) - H(fx_{i_0} \cdots x_{i_{q-1}})$.*

The influence of the variables $x_{i_0}, \ldots, x_{i_{q-1}}$ on the function f is measured by $I(f; x_{i_0} \cdots x_{i_{q-1}})$.

4 Evaluation of Entropies and Decision Trees

Decision trees are used in this section for organizing the computation of various types of entropies and mutual information estimates. We give algorithms that compute the conditional entropy $H(f|x_{i_0} \cdots x_{i_{q-1}})$ as average uncertainty of a function f given values of input variables $x_{i_0}, \ldots, x_{i_{q-1}}$ and the mutual information $I(f; x_{i_0} \cdots x_{i_{q-1}})$ as a measure of the influence of the variables $x_{i_0} \cdots x_{i_{q-1}}$ on f. These quantities are given by

$$\begin{aligned} &H(f|x_{i_0} \cdots x_{i_{q-1}}) \\ &= H(fx_{i_0} \cdots x_{i_{q-1}}) - H(x_{i_0} \cdots x_{i_{q-1}}) \end{aligned} \quad (2)$$

and

$$\begin{aligned} &I(f; x_{i_0} \cdots x_{i_{q-1}}) \\ &= H(f) + H(x_{i_0} \cdots x_{i_{q-1}}) - H(fx_{i_0} \cdots x_{i_{q-1}}). \end{aligned} \quad (3)$$

We attach two types of labelled ordered trees to any logical function.

Definition 4.1 *Let $f : E^n_m \longrightarrow E_m$ be an n-ary, m-valued function and let $q = (x_{i_0}, \ldots, x_{i_{k-1}})$ be a sequence of k distinct variables. The* A-*decision tree of f and q is an m-ary $\mathcal{P}(Dom\,(f))$-ordered tree $T^{f,q}$ such that $T^{f,q}(a_0 \cdots a_{l-1}) = \{u \in Dom\,(f) \mid u[x_{i_0} \cdots x_{i_{l-1}}] = a_0 \cdots a_{l-1}\}$, for every sequence $a_0 \cdots a_{l-1} \in Dom\,(f)$.*

The B-decision tree *of f and q is an m-ary $\mathcal{P}(Dom\,(f))$-ordered tree $S^{f,q}$ such that $S^{f,q}(ba_0 \cdots a_{l-1}) = \{u \in Dom\,(f) \mid f(u) = b$ and $u[x_{i_0} \cdots x_{i_{l-1}}] = a_0 \cdots a_{l-1}\}$, for every sequence $a_0 \cdots a_{l-1} \in Dom\,(f)$ and $b \in ran(f)$.*

Example 4.2 *Consider the function $f : E^3_2 \longrightarrow E_2$ defined by the table:*

	x_0	x_1	x_2	$f(x_0, x_1, x_2)$
u_0	0	0	0	0
u_1	0	0	1	0
u_3	0	1	1	1
u_4	1	0	0	1
u_5	1	0	1	1

The A-decision tree $T^{fx_0x_1x_2}$ is given in Figure 1, while the B-decision tree $S^{fx_0x_1x_2}$ is given in Figure 2.

Using the A-decision tree $T^{f,q}$ and the B-decision tree $S^{f,q}$ we can calculate the entropies $H(r)$ and $H(fr)$, respectively, for any prefix r of q, including q. Note that the j-level of $T^{f,q}$ defines a partition of Dom (f); a block of this partition contains all tuples u such that $u[x_{i_0} \cdots x_{i_{j-1}}] = a_0 \cdots a_{j-1}$.

The total numerical function $R_m : \{1, \ldots, m\} \longrightarrow \mathbb{R}$ is given by

$$R_m(k) = -(k/m)\log_2(k/m)$$

for $1 \leq k \leq m$. For example, R_5 is defined by the table

k	1	2	3	4	5
$R_5(k)$	0.464	0.529	0.442	0.258	0

Now, the entropies $H(r)$ and $H(fr)$ can be expressed using the A- and B-decision trees T^{fq} and S^{fq} for any prefix of the sequence q of variables as $H(r) = -\sum\{R_m(|T^{f,q}(a_0 \cdots a_{|r|-1})|) \mid a_0 \cdots a_{|r|-1} \in$ Dom $(f)\}$,
and $H(fr) = -\sum\{R_m(|S^{f,q}(ba_0 \cdots a_{|r|-1})|) \mid a_0 \cdots a_{|r|-1} \in$ Dom (f) and $f(a_0, \ldots, a_{|q|-1}) = b\}$.

Example 4.3 *The table contained in Figure 4 shows the computation for several entropies associated to the partial function $f : E_2^3 \longrightarrow E_2$ given in Example 4.2. They correspond to decision trees of depth 1. To explore further the entropies associated to f we need to use decision trees of greater depth.*

The entropy $H(x_0x_1)$ for the jointly specified variables x_0, x_1 is $H(x_0x_1) = R_5(2) + R_5(1) + R_5(2) = 1.522$, as it can be seen by inspecting the A-decision tree shown in Figure 1. This allows us to compute the conditional entropies

$$H(x_0|x_1) = H(x_0x_1) - H(x_1) = 0.802,$$
$$H(x_1|x_0) = H(x_0x_1) - H(x_0) = 0.551.$$

Thus, the mutual information $I(x_1; x_2)$ is given by

$$I(x_0; x_1) = H(x_0) - H(x_0|x_1) = 0.971 - 0.802 = 0.169.$$

The same result can be obtained as $I(x_0; x_1) = H(x_1) - H(x_1|x_0) = 0.169$.

The conditional entropies $H(f|x_0)$, $H(x_0|f)$ and the mutual information $I(f; x_0)$ can be computed from the B-decision tree shown in Figure 2. Namely, by examining the nodes located on $S^f_{[2]}$ we obtain:

$$H(fx_0) = R_5(2) + R_5(1) + R_5(2) = 1.522$$

Similar evaluations applied to the decision trees S^{fx_1} and S^{fx_2} shown in Figure 3 yield

$$H(fx_1) = R_5(2) + R_5(2) + R_5(1) = 1.522$$
$$H(fx_2) = R_5(1) + R_5(1) + R_5(1) + R_5(2) = 1.921.$$

Consequently, we obtain

$$\begin{array}{lll} H(f|x_0) = 0.551, & H(x_0|f) = 0.551, & I(f;x_0) = 0.420 \\ H(f|x_1) = 0.802, & H(x_1|f) = 0.551, & I(f;x_1) = 0.169 \\ H(f|x_2) = 0.950, & H(x_2|f) = 0.950, & I(f;x_2) = 0.021 \end{array}$$

These estimates show that the variable x_0 is the one that influences most the value of the function f (since $I(f; x_0) = 0.420$); variable x_0 is weakly connected with the other variables since $I(x_0; x_1) = 0.169$ and $I(x_0; x_2) = 0.021$. In the same time, $I(x_1; x_2) = 0.633$.

5 Future Work

We mention two problems that require further attention. The possiblity of defining the conditional entropy and the mutual information between logical functions (rather than between functions and variables) should be investigated. This should help classify and recognize multivalued functions. Also, an axiomatization of the notion of various types of mutual information seems desirable, along the lines of various axiomatizations for entropy.

6 Acknowledgments

This work was partially supported by KBN (Committee for Scientific Researches), Poland, and by The Fund for Fundamental Research (National Academy of Sciences) and Informatization Fund, Republic of Belarus.

References

[1] Baclawski, K., Simovici, D.: *A Characterization of the Information Content of a Classification*, Information Processing Letters, vol 57, pp. 211-214, 1996.

[2] Chi-Hong Hwang and A.C.-Wu *An Entropy Measure for Power Estimation of Boolean Functions*, Proc. of the ASP-DAC'97–Asia and South Pacific Design Automation Conference, Japan, pp.101–106, 1997.

[3] Ganapathy S., and Rajaraman V., *Information Theory Applied to the Conversion of Decision Tables to Computer Programs*, Comm. of the ACM, vol. 16, pp. 532 - 539, 1973.

[4] Goodman R.M., and Smyth P., *Decision Tree Design from a Communication Theory Standpoint*, IEEE Trans. on Information Theory, vol. 34, no. 5, pp. 979 - 994, 1988.

[5] Hartmann C.R.P., Varshney P.K., Mehrotra K.G., and Gerberich C.L., *Application of Information Theory to the Construction of Efficient Decision Trees*, IEEE Trans. on Inf. Th., vol. IT-28, pp.565-577, 1982.

[6] Kabakcioglu A.M., Varshney P.K., and Hartman C.R.P., *Application of Information Theory to Switching Function Minimization*, IEE Proceedings, Pt E, vol.137, pp.389-393, 1990.

[7] Lioy A., Macli E., Poncino M., and Rossello M., *Accurate Entropy Calculation for Large Logic Circuits Based on Output Clustering*, Proc. Great Lakes Symp. on VLSI, Urbana-Champaign, IL, USA, pp.70-75, 1997.

[8] Lloris-Ruiz A., Gomez-Lopera J.F., and Roman-Roldán R., *Entropic Minimization of Multiple-Valued Functions*, Proc. 23rd Int. Symp. on Multiple-Valued Logic, pp.24-28, 1993.

[9] Lloris A., Gomez J.F., and Roman R., *Using Decision Trees for the Minimization of Multiple-Valued Functions*, Int. J. Electronics, vol.75. no.6, pp.1035-1041, 1993.

[10] Pavlidis T., Swartz J., and Wang J.P., *Fundamentals of Bar Code Information Theory*, IEEE Computer, April, pp.74-86, 1990.

[11] Shmerko V., Cheushev V., and Yanushkevich S. *Classification of Logic Functions based on Information Approach*, Automatics and Telemechanics, Russian Academy of Sciences (in Russian), translated in Automation Remote Control - to be published in Russian and in English.

[12] Simovici D.A. and Reischer C., *On Functional Entropy*, Proc. ISMVL'93, pp.100-104, 1993.

[13] Sollich P., *Minimum Entropy Queries for Linear Students Learning Nonlinear Rules*, Proc. of the 3rd European Symp. on Artificial Neural Networks, pp.217-222, 1995.

[14] Varshney P.K., Hartmann C.R.P., and De Faria J.M., *Application of Information Theory to Sequential Fault Diagnosis*, IEEE Trans. on Computers, vol. C-31, pp.164-170, 1982.

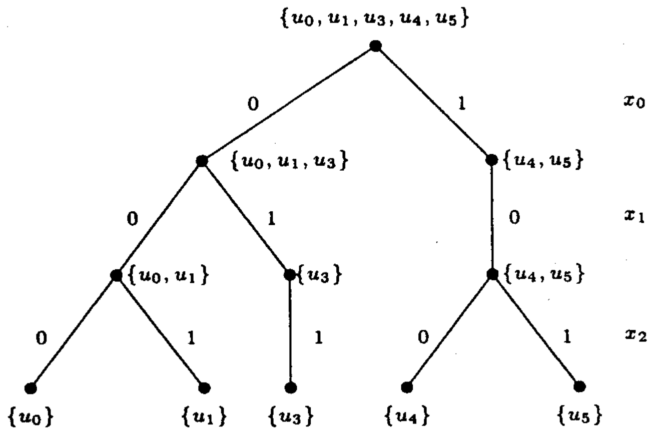

Figure 1: A-decision tree $T^{fx_0x_1x_2}$

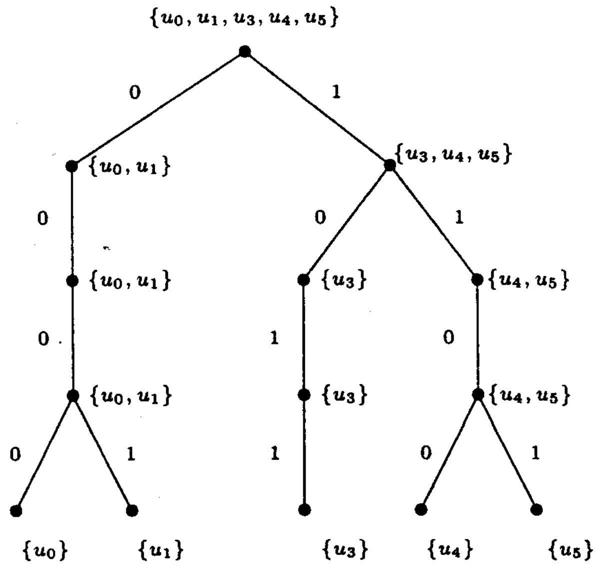

Figure 2: B-decision tree $S^{fx_0x_1x_2}$

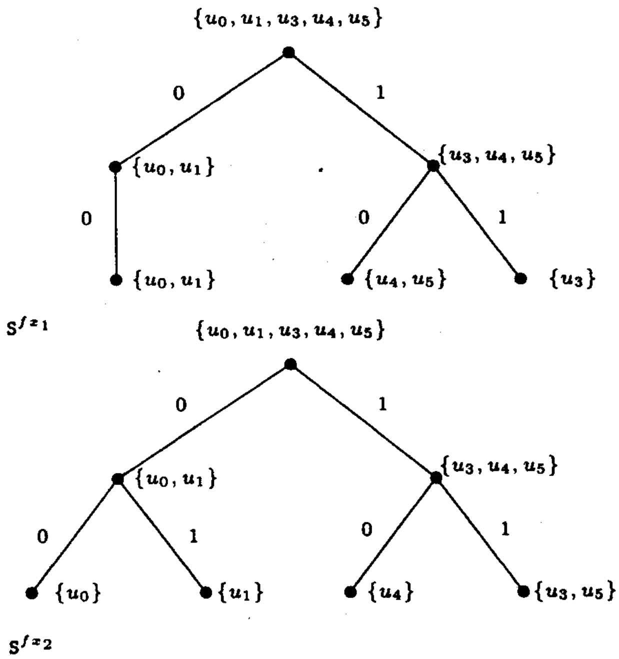

Figure 3: B-decision Trees S^{fx_1} and S^{fx_2}

	x_0		x_1		x_2		f	
Value	0	1	0	1	0	1	0	1
Set of tuples	u_0, u_1, u_3	u_4, u_5	u_0, u_1 u_4, u_5	u_3	u_0, u_4	u_1, u_3, u_5	u_0, u_1	u_3, u_4 u_5
k	3	2	4	1	2	3	2	3
$R_5(k)$	0.442	0.529	0.258	0.464	0.529	0.442	0.529	0.442
Entropy	$H(x_0) = 0.971$		$H(x_1) = 0.729$		$H(x_2) = 0.971$		$H(f) = 0.971$	

Figure 4: Computation of Entropies